Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity

Jiyoung Kim[a], René Villadsen[a], Therese Sørlie[b,c], Louise Fogh[a], Signe Z. Grønlund[a,d], Agla J. Fridriksdottir[a], Irene Kuhn[e], Fritz Rank[f,1], Vera Timmermans Wielenga[f], Hiroko Solvang[b,g], Paul A. W. Edwards[h], Anne-Lise Børresen-Dale[b,i], Lone Rønnov-Jessen[d], Mina J. Bissell[e,2], and Ole William Petersen[a,2]

[a]Department of Cellular and Molecular Medicine, Centre for Biological Disease Analysis and Danish Stem Cell Centre, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark; [b]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello 0310, Oslo, Norway; [c]Cancer Stem Cell Innovation Center, Oslo University Hospital, Norwegian Radium Hospital, 0310 Oslo, Norway; [d]Cell and Developmental Biology, Department of Biology, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark; [e]Life Sciences Division, Berkeley National Laboratory, Berkeley, CA 94720; [f]Department of Pathology, Rigshospitalet, DK-2100 Copenhagen Ø, Denmark; [g]Department of Biostatistics, Institute of Basic Medical Science, University of Oslo, 0317 Oslo, Norway; [h]Department of Pathology and Hutchison/MRC Research Centre, University of Cambridge, Cambridge CB2 0X2, United Kingdom; and [i]K. G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Clinical Medicine, University of Oslo, 0318 Oslo, Norway

Contributed by Mina J. Bissell, February 28, 2012 (sent for review January 10, 2012)

The majority of human breast cancers exhibit luminal epithelial differentiation. However, most aggressive behavior, including invasion and purported cancer stem cell activity, are considered characteristics of basal-like cells. We asked the following questions: Must luminal-like breast cancer cells become basal-like to initiate tumors or to invade? Could luminally differentiated cells within a basally initiated hierarchy also be tumorigenic? To answer these questions, we used rare and mutually exclusive lineage markers to isolate subsets of luminal-like and basal-like cells from human breast tumors. We enriched for populations with or without prominent basal-like traits from individual tumors or single cell cloning from cell lines and recovered cells with a luminal-like phenotype. Tumor cells with basal-like traits mimicked phenotypic and functional behavior associated with stem cells assessed by gene expression, mammosphere formation and lineage markers. Luminal-like cells without basal-like traits, surprisingly, were fully capable of initiating invasive tumors in NOD SCID gamma (NSG) mice. In fact, these phenotypically pure luminal-like cells generated larger and more invasive tumors than their basal-like counterparts. The tumorigenicity and invasive potential of the luminal-like cancer cells relied strongly on the expression of the gene *GCNT1*, which encodes a key glycosyltransferase controlling O-glycan branching. These findings demonstrate that basal-like cells, as defined currently, are not a *requirement* for breast tumor aggressiveness, and that within a single tumor there are multiple "stem-like" cells with tumorigenic potential casting some doubt on the hypothesis of hierarchical or differentiative loss of tumorigenicity.

clonal isolation | prospective | signatures

The two most frequent subtypes of human breast cancer are the luminal and the basal-like, named after their resemblance to the two major lineages in the normal human breast (1). Luminal differentiation independently of estrogen receptor expression, typically includes MUC1 and simple epithelial keratins such as K19 ($MUC1^+/K19^+/ER^{+/-}$), whereas basal-like activity is marked by expression of basal keratins K5, K14, and K17, and/or the transcription factor p63, on an estrogen receptor-low or -negative background ($K5^+/K14^+/K17^+/p63^+/ER^{lo/-}$) (2–4). Despite the tempting extrapolation to luminal vs. basal cells-of-origin for luminal and basal-like breast cancer, respectively, there is increasing evidence to suggest that both subtypes originate from the luminal epithelial lineage (5–7). However, once cancer is established, the only cancer stem cells so far described are essentially basal-like, and the majority of luminal cells within the tumors have been characterized as more differentiated and less malignant than basal cells (8, 9). Adding more complexity is the report that metastatic lesions are enriched in luminal cells (9). A popular explanation for these contradictions is the concept of tumor cell plasticity, i.e., the possibility that differentiated luminal cells must acquire basal-like traits to become malignant (10–14). We set out to examine whether the above explanation always must be true or whether differentiated luminal-like breast cancer cells within a basally initiated hierarchy could be aggressive and stem-like in their own right.

To address these questions, we used two mutually exclusive markers, milk mucin (MM) and CD271, which identify subtypes of cells with either luminal-like differentiation or basal-like activity, respectively. MM was detected by the M18 antibody, which recognizes branched glycans (15), whereas $CD271/p75^{NTR}$ was detected by the ME20.4 antibody (16). We show that frankly differentiated luminal-like cells without acquiring appreciable basal-like traits can be aggressive and invasive when serially transplanted into NOD SCID gamma (NSG) mice or tested for invasiveness in Boyden chambers. Furthermore, that luminal-like cells derived from a stem-like, basal hierarchy cannot only be tumorigenic, but that they can also be more aggressive than their progenitors.

Results

CD271 and MM, Two Distinct Differentiation Markers of Normal Human Breast, Identify Distinct Subsets of Cells in Primary Tumors and Cell Lines. In a search for distinct and rare candidates in both basal and luminal compartments, we used multicolor imaging of normal breast tissue, primary breast carcinomas and established cell lines stained with antibodies against a panel of markers, from which we selected two rare markers, CD271 ($p75^{NTR}$) and MM (branched glycans, here abbreviated MM for 'milk mucin'). These two markers were selected carefully based on their lineage specificity in normal breast, cell surface location, trypsin insensitivity and expression in cultured cells, features that made them ideal for cell sorting and cloning applications (Fig. 1*A Inset*). Our initial staining of 53 biopsies of primary breast tumors revealed that $CD271^+$ cells were present in distinct single cells or small foci in 28 of 53 (53%) biopsies, of which 20 of 28 also contained distinct and nonoverlapping populations of MM^+ cells

Author contributions: J.K., R.V., L.R.-J., M.J.B., and O.W.P. designed research; J.K., R.V., T.S., L.F., S.Z.G., F.R., and V.T.W. performed research; J.K., R.V., T.S., A.J.F., I.K., F.R., V.T.W., H.S., P.A.W.E., A.-L.B.-D., L.R.-J., M.J.B., and O.W.P. analyzed data; and J.K., R.V., I.K., L.R.-J., M.J.B., and O.W.P. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]Deceased January 28th, 2011.

[2]To whom correspondence may be addressed. E-mail: owp@sund.ku.dk or mjbissell@lbl.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1203203109/-/DCSupplemental.

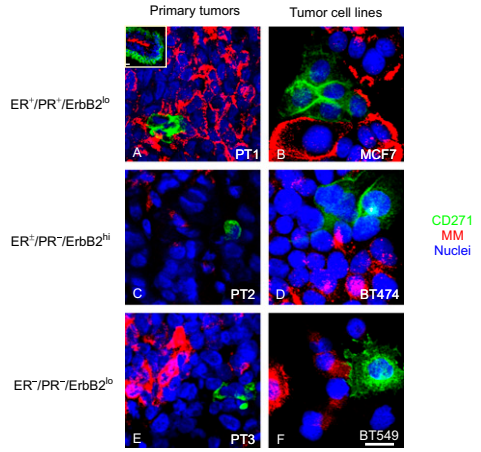

Fig. 1. CD271 and MM, two distinct differentiation markers of normal human breast identify distinct subsets of cells in primary tumors and cell lines. (*A–F*) CD271 and MM staining define separate populations in cryostat sections of tumors (*A*, *C*, and *E*) and in cultures of breast cancer cell lines (*B*, *D*, and *F*) regardless of the subtype. Tumors and established cell lines representative for ER$^+$/PR$^+$/ErbB2^{lo} (*A* and *B*), ER$^{\pm}$/PR$^-$/ErbB2^{hi} (*C* and *D*), and ER$^-$/PR$^-$/ErbB2^{lo} (*E* and *F*) breast tumors were stained with MM (red), CD271 (green), and nuclei (blue). (*A Inset*) Staining of normal breast tissue for comparison. Irrespective of subtype, there was no overlap in staining between MM$^+$ and CD271$^+$. (Scale bar, 25 μm.)

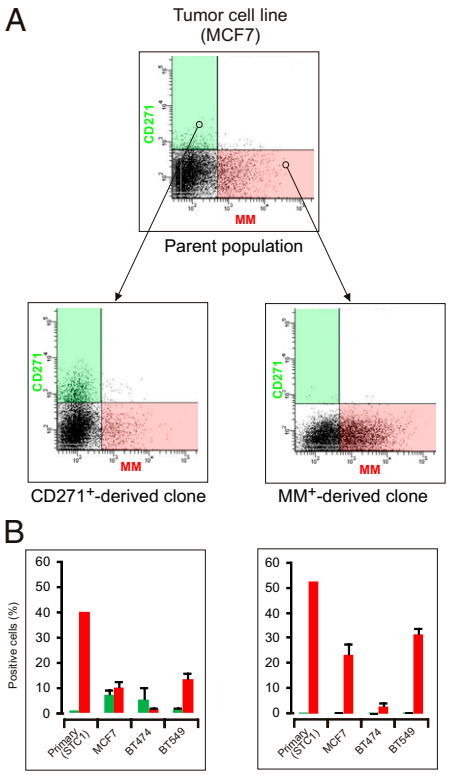

Fig. 2. CD271$^+$-derived clones contain MM$^+$ populations, but freshly isolated MM$^+$ clones do not contain CD271$^+$. (*A*) Representative FACS profiles of a tumor cell line (MCF7) stained with CD271 and MM. Note the L-shaped FACS profile of the parent population including CD271$^+$ (green shade), MM$^+$ (red shade) and double-negative CD271$^-$/MM$^-$ (no shade). Reanalysis of parent-derived clones (encircled) reveal that the CD271$^+$-derived clone regenerates both cell types, whereas the MM$^+$-derived clone does not generate CD271$^+$. (*B*) Frequency of CD271$^+$ (green) and MM$^+$ (red) cells after reanalysis of CD271$^+$- and MM$^+$-derived cells enriched or cloned from primary tumor (STC1) or cell lines. MCF7 and BT474 cell lines could be cloned as single cells. The CD271$^+$ clones consistently gave rise to mixed populations whereas the MM$^+$ clones remained restricted. The data are reported as mean ± SEM (error bar).

(Table S1); 14 of 53 (26%) tumor biopsies exhibited foci of MM$^+$ cells only, and 11 of 53 (21%) showed neither CD271 nor MM staining. The above set of positive populations were present in ER$^+$/PR$^{\pm}$/ErbB2^{lo}, ER$^{\pm}$/PR$^-$/ErbB2^{hi}, and ER$^-$/PR$^-$/ErbB2^{lo} subtypes of breast cancers and the corresponding sample cell lines chosen for further study (MCF7, BT474, and BT549) (Fig. 1). Multicolor imaging revealed that whereas MM was part of a MUC1$^+$/ K19$^+$ compartment, CD271 often costained with p63, K17 and K5 albeit with the latter two exhibiting a more widespread pattern of staining in general, but usually in an essentially ER-and MM-negative background (Fig. S1). Based on these data, we used CD271 as the most stringent criterion for basal-like activity, with p63 as an additional, facultative marker.

CD271$^+$-Derived Clones Contain MM$^+$ Populations, but Freshly Isolated MM$^+$ Clones Do Not Contain CD271$^+$. To determine the relation between the two populations, we combined prospective isolation of cells with single cell cloning. FACS analysis of the MCF7 cell line readily allowed us to separate cells into CD271$^+$ and MM$^+$ gates (Fig. 2*A*). Single-cell cloning efficiency was identical between the gates that allowed delineation of CD271$^+$ and MM$^+$ cells (10–15%), and the expansion of single cells into "clones" required at least a month. That all clones indeed originated from single cells was verified by genomic profiling, and that these underwent essentially the same number of cell divisions before analysis was determined by evaluation of growth curves (Fig. S2 *A* and *B*). Although interclonal variation was recorded, cultures derived from the CD271$^+$ clones recapitulated a pattern of heterogeneity similar to that of the unsorted population, whereas cultures derived from MM$^+$ clones remained essentially negative for CD271 (Fig. 2*A*). The first generation MM$^+$ clones also were negative for other basal-like activity measured including p63, CD44v6, Maspin, and K17 (Fig. S2*C*). Moreover, that CD271$^+$ drove both a basal-like and a luminal-like component, whereas MM$^+$ drove only differentiated luminal-like self-renewal at least in initial passages, was demonstrated with additional cultures derived from prospectively isolated primary tumor-derived cells, as well as clones of MCF7 and BT474, or FACS-sorted cells from BT549 (Fig. 2*B*).

As a proof of principle that pure MM$^+$ luminal-like cells could arise directly from CD271$^+$ cells rather than always being the result of aberrations sustained over length of time the cancer cells have been in culture, we recovered a second round of single cell clones from the newly derived MM$^+$ cells from a freshly isolated CD271$^+$ clone and reanalyzed the resulting progeny for the presence of MM$^+$ and CD271$^+$ cells. Indeed, pure MM$^+$ luminal-like clones emerged in 4 out of 14 clones, confirming that intraclonal unidirectional differentiation does occur even in cancer cell lines. Also, whereas the CD271$^-$/MM$^-$ phenotype was always the most frequent, these were not necessarily similar to one another by the measures we used to assess clonal phenotypes, indicating that there is much additional heterogeneity within this CD271$^-$/MM$^-$ group that remains to be characterized (Fig. S2*D*). The CD271$^-$/MM$^-$ is not currently ripe for further interrogation because we have yet to find tractable and rare markers within this group. Whereas the data in the literature indicate that luminal cells with a basal-like component exhibit lineage marker dynamics in favor of a basal-initiated hierarchy, our data show that this is not always a requirement for clonal expansion.

CD271$^+$ and MM$^+$ Clones Are Distinct Populations by Gene and miRNA Expressions, as Well as by Mammosphere Formation, with only the CD271$^+$ Exhibiting "Stem-Like Activity." We compared the global gene and miRNA expressions of different clones we isolated, using microarray analyses. The cultures derived from CD271$^+$ clones expressed a number of stem cell markers not present in

cultures from MM$^+$ clones (Fig. 3*A*). This was confirmed by quantitative real-time PCR (qRT-PCR) and extended by additional markers (Fig. 3*B*). Several members of the canonical Wnt signaling pathway, shown to be associated with stem cell renewal (17), and a subset of genes that were previously shown to be up-regulated in CD44$^+$ cells (9) were high in the CD271$^+$ clones (Fig. S3). Expression of the miRNAs 205, 221, and 222, which correlate positively with maintenance of mammary epithelial progenitor cells in mice (18, 19) and ER negativity in breast cancer (20), were strongly up-regulated in the CD271$^+$-derived cultures by miRNA microarrays (Fig. 3*C Left*); this result was confirmed by qRT-PCR (Fig. 3*C Right*). We measured progenitor activity using the mammosphere assay, which purportedly measures "stemness" (21): CD271$^+$ cells had enhanced mammosphere-forming capacity compared with MM$^+$ cells (Fig. 3*D*). Thus, for all practical purposes only cells with a basal-like component exhibited stem-like characteristics.

Luminal-Like Cells Within the Hierarchy Are Tumor Initiating and Invasive in the Absence of Basal-Like Traits. To determine how purified differentiated luminal-like cells function in vivo, CD271$^+$ and MM$^+$ cells from both primary tumors and cell lines were xenografted into NSG mice (details of the additional nine primary tumors used is in Table S2). All populations (including CD271$^-$/MM$^-$) formed tumors from relatively low numbers of xenografted cells in NSG mice (Fig. 4*A*). A basal-like-initiated hierarchy was functional in our in vivo assay, because CD271$^+$ cells, but not MM$^+$ cells (with or without basal keratins), were able to recapitulate the heterogeneity of the original, presorted population. This recapitulation was observed irrespective of whether the xenografted cells were derived from primary tumors or tumor cell lines (Fig. 4*B*). Surprisingly, however, we observed that the MM$^+$-derived tumors grew substantially larger than the CD271$^+$-derived tumors and appeared invasive (Fig. 4*C*). We confirmed the invasiveness of each of the populations using the Matrigel-coated transwell filter assay, and the MM$^+$ populations proved by far the most invasive (Fig. 4*D*). We conclude that although there appears to be a differentiation hierarchy, this hierarchy does not confer progressive loss of tumor-initiating capacity and invasiveness. These properties do not diminish with differentiation; if anything, the opposite is true.

To understand the mechanism responsible for the MM$^+$ cells display the robust malignant phenotype, we investigated the role of expression of the branched glycan specific to MM cells. This glycan is detected by MAb M18, an antibody that has previously been identified as a differentiation marker of breast epithelium (22). The primary gene controlling branching of O-glycans in breast epithelium is the core 2 synthase (core 2 β6GlcNAcT1) encoded by GCNT1. We therefore tested whether GCNT1 may play a role in the aggressiveness of the MM$^+$ subpopulation (Fig. 4*E* and Fig. S4*A*). Knockdown of GCNT1 did not interfere with expression of MUC1 as revealed by staining with the antibody 115D8 (Fig. S4*B*). GCNT1 knockdown in cultures derived from a MM$^+$ clone of MCF7 dramatically delayed tumor formation in nude mice to a level comparable to the delay observed for cultures derived from CD271$^+$ clones (Fig. 4*F*); in addition, there was a significant reduction in the invasive capacity as measured by the Matrigel-coated transwell filter assay (arrow in Fig. 4*D*). These data support the surprising conclusion that aggressiveness of MM$^+$ tumors is dependent on specific changes in glycosylation.

CD271$^+$ and MM$^+$ Gene Signatures Predict Poor Clinical Outcome in Breast Cancer. To determine the clinical implications of these results, we analyzed the gene signatures for the CD271$^+$ and MM$^+$ tumors. We selected a set of 1,025 probes that were among

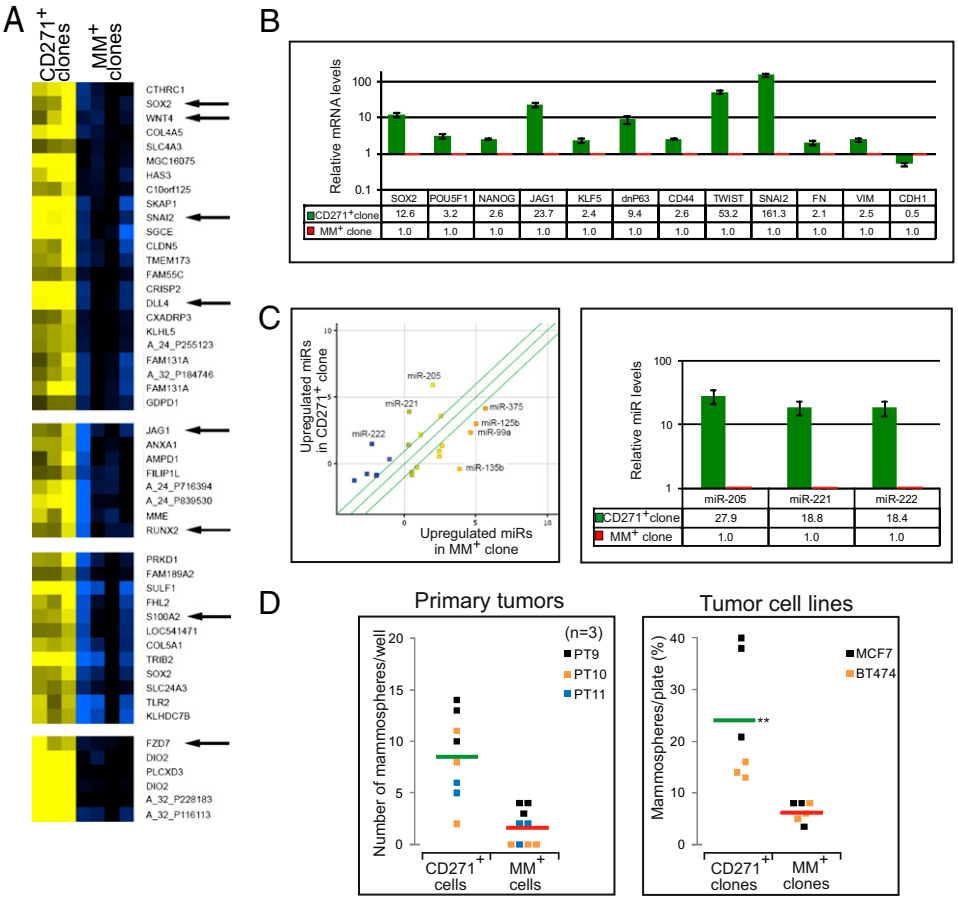

Fig. 3. CD271$^+$ and MM$^+$ clones are distinct populations by gene and miRNA expressions, and mammosphere formation with only the CD271$^+$ exhibiting stem cell properties. (*A*) Heat maps of selective gene sets from clustered data are shown depicting high expression of specific stem-cell makers (arrows) in CD271$^+$ clones compared with MM$^+$ clones of MCF7. Yellow = highest expression; blue = lowest expression; black = average expression; gray = missing data. (*B*) Differential gene expression verified and expanded by qRT-PCR suggests stem-like properties of CD271$^+$ as opposed to MM$^+$ cells. (*C*) miRNAs differentially expressed between CD271$^+$ and MM$^+$ clones of MCF7. Expression was determined by microarray analysis (*Left*) and confirmed by qRT-PCR (*Right*). The blue to green dots indicate that the values are higher in CD271, whereas orange to red indicate that the values are higher in the MM cells. The miRNAs that show at least twofold change between the two populations are presented. Note that miRs 205, 221 and 222 generally associated with mammary progenitor activity were highly expressed in the CD271$^+$ clones. (*D*) CD271$^+$ and MM$^+$ populations have distinct mammosphere-forming capacity. Mammospheres were quantified as numbers of mammospheres per well (tumors) or frequency of wells with mammosphere formation (cell lines). $^{**}P < 0.01$ compared with MM$^+$.

A

Source	Sample	Subpopulation xenografted	Tumor incidence/xenograft		
			No. of cells per xenograft		
			10.000	1.000	100
Primary tumor	PT13	$CD271^+$	2/2	1/1*	3/3*
		MM^+	2/2	4/5*	3/5*
		$CD271^-/MM^-$	2/2	2/3*	3/3*
Short-term cultured	STC1	$CD271^+$			0/2
		MM^+		1/2	0/2
		$CD271^-/MM^-$		0/2	0/2
Short-term cultured	STC2	$CD271^+$	2/2	1/2	1/2
		MM^+	2/2	2/2	1/2
		$CD271^-/MM^-$			
Tumor cell line	MCF7	cloned $CD271^+$	2/2	2/2 (4/4*)	3/3
		cloned MM^+	2/2	2/2 (4/4*)	3/3
		$CD271^-/MM^-$			2/2

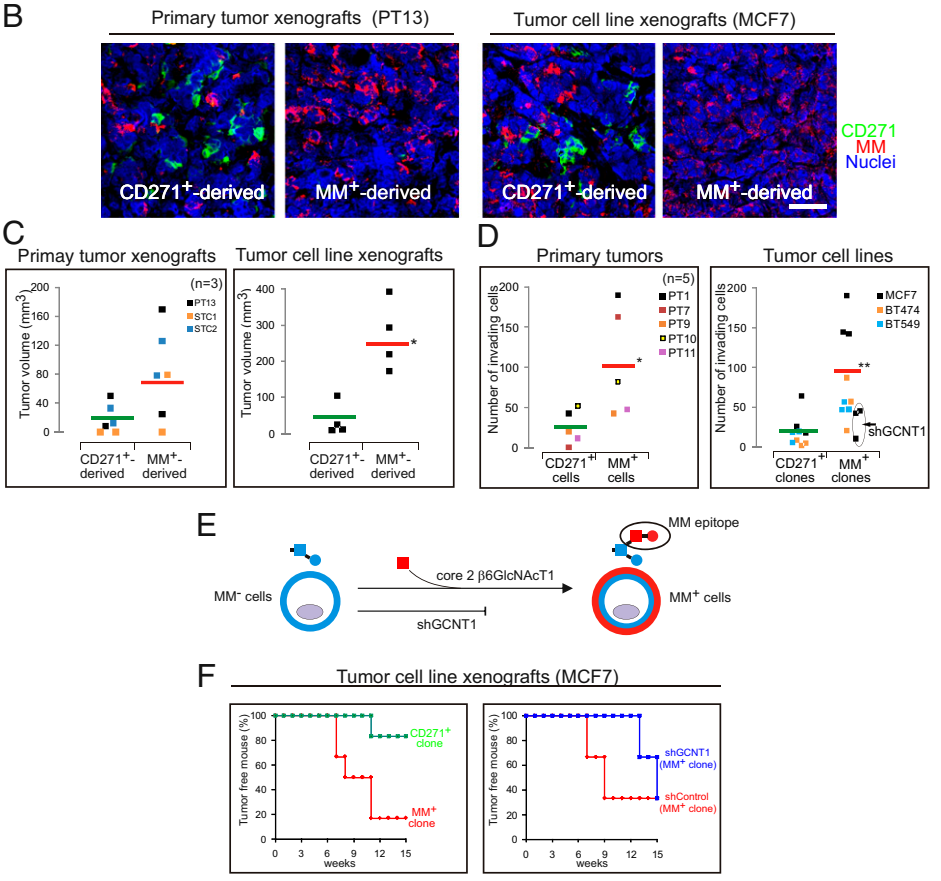

Fig. 4. Luminal-like cells within the hierarchy are tumor initiating and invasive in the absence of basal-like traits. (*A*) Summary of tumor formation in NSG mice. Tumor formation within 10–15 wk after injection into NSG mice of cells from a xenografted tumor (PT13), from short-term cultured samples (STC1, STC2), or from a tumor cell line (MCF7) all separated by flow cytometry into $CD271^+$ (green), MM^+ (red), and $CD271^-/MM^-$ (blue) populations. All cell populations readily formed tumors. Asterisks indicate tumor formation by serial transplantation. (*B*) Multicolor imaging of cryostat sections from $CD271^+$- and MM^+-derived xenografts stained with CD271 (green), MM (red), and nuclei (blue). Note that the hierarchical organization of tumor phenotype is maintained also in vivo. (Scale bar, 50 μm.) (*C*) Tumor volume of $CD271^+$- and MM^+-derived xenografts (second passage PT13 and STC1, first passage SCT2, second passage MCF7 clones) as measured 10–15 wk after injection of 10^3–10^4 cells. For STC1 the $CD271^+$ cells were substituted with MM^-/EpCAM$^+$ cells due to lack of sufficient numbers of bona fide $CD271^+$ cells. Dot plots indicate individual tumor volumes (Mean indicated by solid line). $^*P < 0.05$ compared with $CD271^+$. (*D*) Cells sorted by flow cytometry into $CD271^+$ and MM^+ from primary tumors (1×10^4 cells; $n = 5$) and tumor cell lines (1×10^5 cells of MCF7 clones or BT474 clones; 5×10^4 cells of FACS-sorted BT549; $n = 3$ each cell line) were incubated in a modified Boyden chamber assay with thinly layered Matrigel-coated Transwell chambers. For some of primary tumors (PT1, PT7, PT9) or for BT549, $CD271^+$ cells were replaced with MM^-/EpCAM$^+$ or $MM^-/CD271^-$ due to insufficient numbers of $CD271^+$ cells. Arrow indicates MM^+ cell invasion in vitro after GCNT1 silencing. Silenced cells were less invasive than the control cells. $^*P < 0.05$; $^{**}P < 0.01$ compared with $CD271^+$. (*E*) Schematic representation of a *GCNT1* silencing strategy in MM^+ cells. The core 2 β6GlcNAcT1 encoded by GCNT1 controls branching of O-glycans, here assumed to represent the MM epitope. Short hairpin RNA targeting of GCNT1 (shGCNT1) inhibits expression of the MM epitope. (*F*) Kaplan–Meier graphs depict the incidence of palpable tumor development over time after s.c. injection of 10^7 cells from the $CD271^+$ or MM^+ clones of MCF7 in nude mice, showing that GCNT1 silencing ($n = 3$) inhibits tumor formation of MM^+ clone to the level of $CD271^+$ clone in nude mice. MM^+ cells are much more aggressively tumorigenic in nude mice. Five out of six injections developed tumors with MM^+ cells.

the top genes identified by differential expression between $CD271^+$ and MM^+ clones using significance analysis of microarrays (SAM). (Fig. 5*A*; probes and genes listed in Dataset S1). In three independent datasets, we found that both $CD271^+$ or MM^+ gene signatures predicted poor, relapse-free survival from breast cancer ($P < 0.05$) (Fig. 5*B* and Fig. S5 *A* and *B*). Remarkably, the relapse-free survival curves are almost identical whether one uses the $CD271^+$ or the MM^+ signatures despite the fact that the two signatures are not related. As control, a specific gene set of highly expressed genes in parental MCF7, different from the $CD271^+$ and MM^+ signatures, was not predictive of outcome in the same cohorts. Thus, both CD271- and MM-derived gene expression signatures show prognostic value.

Discussion

We have prospectively isolated two distinct populations of malignant cells, one basal-like and one luminal-like, forming a differentiation hierarchy in primary breast tumors as well as in diverse breast cancer cell lines. Both the population with the basal-like and the luminal-like markers are tumor-initiating and invasive by all of the criteria measured. These data could be interpreted either in light of the cancer stem cell (CSC) model or the clonal evolution model (for review, see ref. 23). Our finding of a stem-like subpopulation, which by all measures appears to recapitulate the heterogeneity of the original population, would be expected within a CSC framework. Such prediction is a hallmark of the CSC hypothesis (8). However, our results differ dramatically from the CSC model in that all of the populations tested within the differentiation hierarchy are tumor initiating at relatively low cell numbers in the more permissive NSG mice. One explanation for this apparent contradiction could be the possibility that differentiated cancer cells potentially convert into tumor-initiating CSCs in a reversible manner (11). Breast CSCs are thought to be

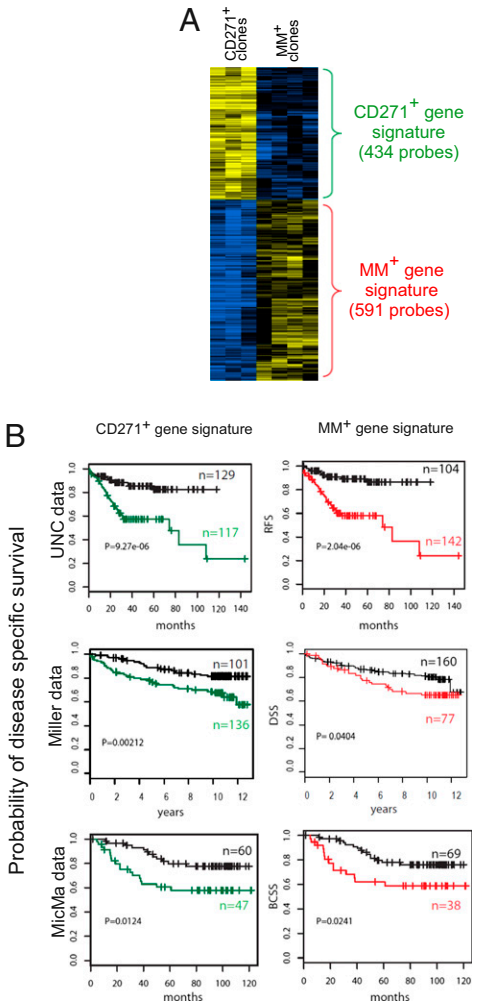

Fig. 5. CD271$^+$ and MM$^+$ gene signatures predict poor clinical outcome in breast cancer. (*A*) Scaled-down representation of the 1025 probe cluster, illustrating statistically differentially expressed genes between four MM$^+$ and three CD271$^+$ clones of MCF7. 591 probes that were higher expressed in MM$^+$ cells compared with CD271$^+$ cells, are referred to as the 'MM$^+$ gene signature', whereas 434 probes that were higher expressed in CD271$^+$ cells compared with MM$^+$ cells are referred to as 'CD271$^+$ gene signature'. (*B*) Disease-specific survival for patients separated in two groups based on expression of CD271$^+$ or MM$^+$ gene signature. Expression of the gene signatures was analyzed in three independent breast cancer microarray datasets: UNC dataset with 243 patients, Miller dataset with 237 patients, and MicMa dataset with 107 patients. Kaplan–Meier curves and log rank test show that both gene signatures are able to identify patients with higher rates of breast cancer recurrence.

generally basal-like (9, 14), implying that tumor formation would have to be initiated by cells with basal-like activity.

By applying two narrowly expressed, mutually exclusive markers, MM and CD271, not used previously for breast CSC analysis, here we provide evidence compatible with a third, "fusion" model. CD271 has been described as a unique marker of melanoma stem cells in humans (24). However, its first description in breast attributed it as a marker of myoepithelial cells (16). More recent studies show that its overexpression strongly increases resistance to anti-tumoral TRAIL treatment (25). The M18 antibody was shown to recognize branched glycans (15) and to stain apical membranes of luminal epithelial cells in the normal human breast (22), but further specificity of its affinity remained to be investigated. Here we identify a linear connection between cells expressing these two nonoverlapping markers suggestive of at least a rudimentary differentiation lineage hierarchy in breast cancers—and surprisingly also in breast cancer cell lines. We show, however, that differentiation within this hierarchy not only does not decrease tumor-initiating capacity as predicted by the CSC model, but instead, the expanded sets of derivative cells appear to function as additional candidate cancer-initiating cells driving aggressive tumors. However, these cells are distinct from canonical stem cells in that they neither have undergone a detectable epithelial-to-mesenchymal transition, nor have they regressed to any particular dedifferentiated or basal-like state. These data require a revision to the existing hypotheses of breast cancer progression that can account for the observed contributions of multiple cell populations to tumor initiation.

Another unexpected finding in the present study is that basal-like- or luminal-like- differentiation within subpopulations appears to not be limited to any particular subtype of breast cancer. In fact, we found rare basal-like cells even in bona fide primary luminal breast tumors or more surprisingly in breast cancer cell lines, such as MCF7, and found MM$^+$ luminally differentiated cells in bona fide primary basal tumors or basal tumor cell lines such as BT549. This could be interpreted either in favor of the classical view of cancer as a caricature of normal tissue renewal (26), or as transformation-induced basal/luminal-like reprogramming (27). The first interpretation would have to rely on the assumption that breast cancer originates from normal stem cells. Candidate human breast stem cells have been reported repeatedly to be basal-primed by several of the markers used in the present study to characterize basal-like activity (9, 28). It was shown recently in mice that adult tissue homeostasis does not rely on classical stem cells, but rather on self-duplication of lineage-restricted progenitors (29). However, in contrast to mice, humans contain a substantial amount of K5$^+$/K14$^+$/K17$^+$ luminal breast epithelial cells, which is why basal-like in the human breast is better defined by the concomitant expression of p63 and CD271. Nevertheless, if breast epithelial multipotency were limited to fetal life also in humans, and basal-like activity in its most stringent CD271$^+$ sense reflected cell-of-origin in cancer, it would imply that the earliest seeds of human breast cancer must exist in utero. Although increased susceptibility to breast cancer has been linked to the intrauterine origin (30), the preponderance of data that suggest a luminal origin of most breast cancers favors the second option for generation of basal-like activity in breast cancers, i.e., postnatal or postpuberty reprogramming (5, 6). This possibility includes that luminal cells should occasionally acquire a basal-like phenotype upon malignant transformation. We have shown previously that under certain culture conditions a small minority can convert to basal-like cells (31). With the recent demonstration of in vivo self-duplication in the mouse mammary gland in mind (29), it may be that culture conditions may unmask latent (or cryptic) potentials in luminal cells. Such distinctive expression of gene expression potentials between cells in vivo and cells in culture was summarized long ago and continues to be reported (32, 33). In fact, it all comes down to microenvironmental cues and context because, even in vivo, mouse mammary gland luminal progenitors generate basal cells if exposed to a cleared fat pad (34, 35).

Finally, the most intriguing finding of our investigation, apart from the fact that multiple cell types are tumor initiating, is the observation that both invasion and tumorigenicity of luminal-like breast cancer and breast cancer cell lines depended on GCNT1 expression, which is required for elaboration of the MM-specific branched glycan recognized by antibody M18 (15). There is increasing appreciation that glycans play an important role in malignant behavior. Indeed, GCNT2 has been associated with breast cancer metastasis, and blocking its expression was shown to abrogate migration and invasion (36).

In conclusion, our data implicate multiple types of tumor-initiating cells in breast cancers camouflaged in different phenotypic cloaks, including differentiated luminal-like cells. This particular finding is of profound clinical importance and strongly suggests the need for combinatorial therapies targeting multiple

cell types in our search to prevent cancer recurrence or curb tumor aggressiveness while sparing the host.

Materials and Methods

Human Tissue Samples and Mice. Human tissue included fifteen reduction mammoplasty samples and eighty-one mastectomy biopsies. Mice included BALB/cA nude mice as well as NSG mice. See *SI Materials and Methods*.

Human Breast Cancer Cell Lines. Cell lines included three established cell lines (MCF7, BT474, and BT549) and two primary derived cell lines defined as cultures expanded within the limits of a normal finite life span (HMT-3909 and L56Br-C1). See *SI Materials and Methods*.

Cell Isolation, Staining, and Sorting. Tumor tissue was enzymatically digested by collagenase and trypsin/EDTA into single cells before FACS analysis. Single cell clones were generated from established cell lines, and >50 clones were reanalyzed for multipotency. See *SI Materials and Methods*.

Antibodies and Immunostaining. Cryostat sections of normal or tumor tissue as well as cultured cells were stained by immunoperoxidase or immunofluorescence. Tumors were classified as $ER^+/PR^+/ErbB2^{lo}$, $ER^{\pm}/PR^-/ErbB2^{hi}$, or $ER^-/PR^-/ErbB2^{lo}$ phenotype based on staining with ER, PR, ErbB2, K5, and K17. See *SI Materials and Methods*.

In Vitro Cellular Assays. See *SI Materials and Methods*.

In Vivo Transplantation. Limited dilution of MM^+ or $CD271^+$ MCF7 clones in a suspension of 50% Matrigel (BD Biosciences, cat. no. 356231) in DME/F12 medium with 10% (vol/vol) FCS was performed in NSG mice for 8–10 wk. $MM^-/CD271^-$ cells of MCF7 were tested by two inoculations of 10^2 cells. For serial transplantation, MM^+ or $CD271^+$ cells isolated by FACS from an initial inoculation of 10^2 MM^+ or $CD271^+$ MCF7 clones were retransplanted in NSG mice with 10^3 cells. Tumor growth was monitored for 10 wk, and tumor volume was measured at the time of sacrifice and calculated by the ellipsoid volume (37). For serial transplantation, cells were isolated by FACS, transplanted into NSG mice and monitored for up to 15 wk. See *SI Materials and Methods*.

RNA Isolation and mRNA/miRNA Expression Analysis. See *SI Materials and Methods*.

Microarray Experiments. See *SI Materials and Methods*.

GCNT1 Silencing. See *SI Materials and Methods*.

Data Analysis and Statistics. Genes significantly differentially expressed between MM^+ and $CD271^+$ cells were identified using two-class, unpaired SAM and 1025 genes were confirmed by DEDS. Survival analysis was performed by Kaplan–Meier analysis with log-rank test. Nonparametric Mann–Whitney tests were performed to measure the differences in mammosphere-forming capacity, tumor formation, and invasiveness between $CD271^+$ and MM^+ cells. See *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Tove Marianne Lund, Lena Kristensen, Hilde Johnsen, Mimi Birkelund, and Margit Bæksted for expert technical assistance and for animal studies and Drs. Alexander Borowsky, Henrik Clausen, Joe W. Gray, Curt Hines, Mark LaBarge, and Kornelia Polyak for critically reading various versions of the manuscript. We also thank Søllerød Privathospital and Københavns Privathospital for the breast biopsy materials and Dr. Åke Borg for L56Br-C1 cells. The work from our laboratories is supported by Danish Cancer Society Grants DP07063 and R20-A1149-10-S2; European Commission Contract LSHC-CT-2006-037632 (to the European Cancer Stem Cell Consortium); Danish Agency for Science and Technology Innovation 2107-05-0006 (to DAN-ED: Endocrine disruptors in food and environment), 10-092798 (to DANSTEM), and 08-045450 (to Danish-Japanese Cooperative Research); the Dansk Kræftforskningsfond; the Lundbeck Foundation; the Novo Nordisk Foundation; the Simon Spies Foundation; The John and Birthe Meyer Foundation; and Fru Astrid Thaysens Legat for Lægevidenskabelig Grundforskning. The work of the A.-L.B.D. and T.S. laboratories is supported by grants from the Research Council of Norway and the Norwegian Cancer Society. Work from M.J.B.'s laboratory is supported by grants from the US Department of Energy, Office of Biological and Environmental Research and Low Dose Radiation Program Contract DE-AC02-05CH1123; by National Cancer Institute Awards R37CA064786, U54CA126552, U54CA112970, U01CA143233, and U54CA143836 (Bay Area Physical Sciences–Oncology Center, University of California, Berkeley, CA); and by US Department of Defense W81XWH0810736.

1. Sorlie T, et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 100:8418–8423.
2. Abd El-Rehim DM, et al. (2005) High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int J Cancer* 116:340–350.
3. Badve S, et al. (2011) Basal-like and triple-negative breast cancers: A critical review with an emphasis on the implications for pathologists and oncologists. *Mod Pathol* 24:157–167.
4. Villadsen R, et al. (2007) Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol* 177:87–101.
5. Molyneux G, et al. (2010) BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 7:403–417.
6. Proia TA, et al. (2011) Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* 8:149–163.
7. Lim E, et al. (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* 15:907–913.
8. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA* 100:3983–3988.
9. Shipitsin M, et al. (2007) Molecular definition of breast tumor heterogeneity. *Cancer Cell* 11:259–273.
10. Chaffer CL, et al. (2011) Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc Natl Acad Sci USA* 108:7950–7955.
11. Gupta PB, et al. (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* 146:633–644.
12. Horwitz KB, Dye WW, Harrell JC, Kabos P, Sartorius CA (2008) Rare steroid receptor-negative basal-like tumorigenic cells in luminal subtype human breast cancer xenografts. *Proc Natl Acad Sci USA* 105:5774–5779.
13. Meyer MJ, et al. (2009) Dynamic regulation of CD24 and the invasive, CD44posCD24-neg phenotype in breast cancer cell lines. *Breast Cancer Res* 11:R82.
14. Kabos P, et al. (2011) Cytokeratin 5 positive cells represent a steroid receptor negative and therapy resistant subpopulation in luminal breast cancers. *Breast Cancer Res Treat* 128:45–55.
15. Gooi HC, et al. (1983) Two mouse hybridoma antibodies against human milk-fat globules recognise the I(Ma) antigenic determinant beta-D-Galp-(1 leads to 4)-beta-D-GlcpNAc-(1 leads to 6). *Carbohydr Res* 120:293–302.
16. Popnikolov NK, Cavone SM, Schultz PM, Garcia FU (2005) Diagnostic utility of p75 neurotrophin receptor (p75NTR) as a marker of breast myoepithelial cells. *Mod Pathol* 18:1535–1541.
17. Zeng YA, Nusse R (2010) Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* 6:568–577.
18. Ibarra I, Erlich Y, Muthuswamy SK, Sachidanandam R, Hannon GJ (2007) A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. *Genes Dev* 21:3238–3243.
19. Greene SB, Gunaratne PH, Hammond SM, Rosen JM (2010) A putative role for microRNA-205 in mammary epithelial cell progenitors. *J Cell Sci* 123:606–618.
20. Zhao JJ, et al. (2008) MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *J Biol Chem* 283: 31079–31086.
21. Dontu G, et al. (2003) In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev* 17:1253–1270.
22. Edwards PA, Brooks IM (1984) Antigenic subsets of human breast epithelial cells distinguished by monoclonal antibodies. *J Histochem Cytochem* 32:531–537.
23. Marusyk A, Polyak K (2010) Tumor heterogeneity: Causes and consequences. *Biochim Biophys Acta* 1805:105–117.
24. Boiko AD, et al. (2010) Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. *Nature* 466:133–137.
25. Verbeke S, et al. (2010) Overexpression of p75(NTR) increases survival of breast cancer cells through p21(waf1). *Cell Signal* 22:1864–1873.
26. Pierce GB, Speers WC (1988) Tumors as caricatures of the process of tissue renewal: prospects for therapy by directing differentiation. *Cancer Res* 48:1996–2004.
27. Visvader JE (2011) Cells of origin in cancer. *Nature* 469:314–322.
28. Gudjonsson T, et al. (2002) Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev* 16:693–706.
29. Van Keymeulen A, et al. (2011) Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479:189–193.
30. Moral R, et al. (2011) In utero exposure to butyl benzyl phthalate induces modifications in the morphology and the gene expression profile of the mammary gland: An experimental study in rats. *Environ Health* 10:5.
31. Petersen OW, et al. (2003) Epithelial to mesenchymal transition in human breast cancer can provide a nonmalignant stroma. *Am J Pathol* 162:391–402.
32. Bissell MJ (1981) The differentiated state of normal and malignant cells or how to define a "normal" cell in culture. *Int Rev Cytol* 70:27–100.
33. Joseph NM, et al. (2011) Enteric glia are multipotent in culture but primarily form glia in the adult rodent gut. *J Clin Invest* 121:3398–3411.
34. Sleeman KE, et al. (2007) Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. *J Cell Biol* 176:19–26.
35. Bruno RD, Boulanger CA, Smith GH (2012) Notch-induced mammary tumorigenesis does not involve the lobule-limited epithelial progenitor. *Oncogene* 31:60–67.
36. Zhang H, et al. (2011) Engagement of I-branching beta-1, 6-N-acetylglucosaminyltransferase 2 in breast cancer metastasis and TGF-beta signaling. *Cancer Res* 71:4846–4856.
37. Tomayko MM, Reynolds CP (1989) Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother Pharmacol* 24:148–154.

Supporting Information

Kim et al. 10.1073/pnas.1203203109

SI Materials and Methods

Human Tissue Samples and Mice. Normal breast tissue ($n = 15$) was obtained from patients undergoing reduction mammoplasties for cosmetic reasons at Søllerød Privathospital and Københavns Privathospital from consenting individuals, and tumor specimens ($n = 81$) were obtained from mastectomies performed at the State University Hospital, Copenhagen, with the approval from the Regional Scientific Ethical Committees for Copenhagen and Frederiksberg (KF) (11)263995 and for Region Hovedstaden H-2–2010-051. Athymic nude mice (BALB/cA nude mice, Taconic) and NOD scid gamma (NSG) mice (The Jackson Laboratory) were maintained in the University's animal facility according to institutional guidelines. All mouse experiments were conducted in accordance with procedures approved by the Animal Experiments Inspectorate.

Human Breast Cancer Cell Lines and Primary Tumor-Derived Cells. MCF7 cells were obtained and cultured as described (1), and BT474, and BT549 were purchased from ATCC. To avoid contaminating nonmalignant cells primary tumor-derived cultures were expanded for a while in short-term culture (STC) until purity with respect to tumor cells. STC defined as cultures expanded within the limits of a normal finite life span included HMT-3909 (STC1, passage 10) isolated as described (2), and L56Br-C1 (3) (STC2, passage 50), which was a kind gift from Professor Åke Borg (Lund University). Unless otherwise indicated, all media and supplements were purchased from Gibco. MCF7 cells were cultured in DME/F12 medium supplemented with 2mM glutamine, 50mg/mL gentamycin (Biological Industries), 5 μg/mL insulin (Sigma), and 5% FBS. BT474 cells were cultured in DMEM 1965 medium with glutamine, gentamycin, and 10% FBS, and BT549 cells were cultured in RPMI medium 1640 with glutamine, gentamycin, 1 mM sodium pyruvate, 2.5 mg/mL glucose and 10% FBS. HMT-3909 cells were cultured in CDM3 medium as described (4) with 20% FBS and L56Br-C1 cells were cultured in RPMI medium 1640 with glutamine, gentamycin, 7 NEA, 1 mM sodium pyruvate, 10 μg/mL insulin, and 10% FBS. All cells were cultured at 37 °C in a humidified atmosphere with 5% CO_2.

Cell Isolation, Staining, and Sorting. Tumor tissue was minced with scalpels and enzymatically digested by overnight incubation, except the xenografts by two hour incubation, in collagenase Type IV (900 units/mL, Worthington Biochemicals) in DMEM/F12 medium supplemented with 2 mM glutamine (Gibco), 50 mg/mL gentamycin and 5% FBS, at 37 °C on a rotary shaker. Tumor cell aggregates were further dissociated into single cells with 0.25% trypsin/1mM EDTA for 10 min and sieve filtration through a 100-μm cell strainer. Short-term cultured tumor cells and cell lines were first trypsinized, filtered through a 10-μm cell strainer (BD Biosciences), before analysis with a FACSAria cell sorter (BD Biosciences) as described (4), using the M18 (1:100, LICR-LON-M18, provided by Professor Paul A.W. Edwards), anti-CD271 (1:100, ME20.4, Abcam), anti-CD271-APC (1:10, ME20.4, Cedarlane Laboratories), anti-EpCAM (1:100, VU1D9, Novocastra), anti-EpCAM-PercpCy5.5 (1:20, 9C4, Biosite), or anti-H2KD-PE (1:20, SF1-1.1, BD biosciences) antibodies with the isotype-specific secondary fluorescent antibodies (Alexafluor, Invitrogen; BD Pharmingen, BD Biosicences), if antibodies were not conjugated. To generate $CD271^+$ and MM^+ clones from breast cancer cell lines, single cells from $CD271^+$ and MM^+ gates were sorted separately into individual wells of 96-well plates (Nunc) and expanded. More than sixty clones were reanalyzed. Generally clones or enriched 'primary' tumor cells (STC1) were reanalyzed for e.g., multipotency after 4 wk, at which time they had undergone ~11 population doublings and contained ~2,000 cells.

Antibodies and Immunostaining. Cryostat sections of normal or tumor tissue as well as cultured cells were stained by immunoperoxidase or immunofluorescence as described (4, 5). For initial subtyping of biopsies, anti-ER (1:100, 1D5, Dako), anti-PR (1:100, Pgr636, Dako), anti-ErbB2 (1:200, TAB250, Zymed), anti-K17 (1:100, E3, Dako), anti-K5 (1:250, XM26, Novocastra), and anti-p63 (1:25, 7JUL, Novocastra) were used for immunoperoxidase. Nuclei were counterstained with hematoxylin (Sigma). For fluorescence, sections were stained with M18 (1:25), anti-CD271 (1:25), anti-ΔNp63 (1:25, BioLegend), anti-K17 (1:10), anti-K5 (1:100), or anti-ER (1:25, Sp1, Labvision). Nuclei were counterstained with propidium iodide (1 μg/mL, Invitrogen). $CD271^+$ and MM^+ clones were stained by immunoperoxidase with p63 and K17 in addition to anti-CD44v6 (1:200, VFF7, Abcam) and anti-Maspin (1:200, G167-70, BD biosciences). Tumors were classified as $ER^+/PR^+/ErbB2^{lo}$ (luminal), $ER^{\pm}/PR^-/ErbB2^{hi}$ (ErbB2), or $ER^-/PR^-/ErbB2^{lo}$ (basal-like or triple-negative) phenotype based on staining with ER, PR, ErbB2, K5 and K17.

In Vitro Cellular Assays. To quantify cell invasion, 1×10^5 cells of MCF7 or BT474 clones, 5×10^4 cells of FACS-sorted BT549 ($n = 3$ of each cell line) and 1×10^4 primary cells ($n = 5$ biopsies) were plated in transmembrane chambers (8-μm pore size, 24-well plate format) coated with thinly layered Matrigel (Millipore) and supplemented with 0.5% FBS for cell lines and 5% FBS in CDM3 for primary cells. Generally, the lower chambers were supplemented with the growth media for cell lines and CDM3 medium with 20% FBS for primary tumor cells, and incubated for 4~5 d except that BT549 was incubated for 2 d. To support migratory behavior, the BT474 cell growth medium was supplemented with 10 ng/mL hepatocyte growth factor (Sigma) and 1 μM progesterone (Sigma) 1 wk before performing the invasion assay. Invaded cells stained with 0.4% crystal violet in 50% ethanol were counted by light microscopy at 20× magnification by two observers using an ocular grid.

Single-cell mammosphere assay was performed as described (5). MCF7 cells were cultured in MEGM medium with supplements and BT474 cells were cultured in regular growth medium in 96-well plates. The number of spheres was counted after 2 wk in culture. Freshly sorted $CD271^+/EpCAM^+$ or $MM^+/EpCAM^+$ cells from primary breast tumors were cultured in low attachment 24-well plates (2,000 cells per well in triplicate) in MEGM medium, and the number of spheres per well that were bigger than 80 μm was quantified after 2 wk.

In Vivo Transplantation. Limited dilution of MM^+ or $CD271^+$ MCF7 clones in a suspension of 50% Matrigel was performed in 8-wk-old female NSG mice with two to four inoculations of 10^6–10^2 cells for 8 wk. Drinking water of mice receiving estrogen receptor positive cells was supplemented with 0.67 μg/mL 17β-estradiol (Sigma, Catalogue #E2758; ref. 6) from 1 wk in advance of inoculation until killed. Fresh estradiol-supplemented water was provided twice a week. $MM^-/CD271^-$ cells of MCF7 were tested by two inoculations of 10^2 cells. For serial transplantation, MM^+ or $CD271^+$ cells isolated by FACS from an initial inoculation of 10^2 MM^+ or $CD271^+$ MCF7 clones, were retransplanted in NSG mice with 10^3 or 10^2 cells. Tumor growth was monitored for 10 wk, tumor volume was measured at the

time of sacrifice, and calculated by the ellipsoid volume (7). For serial transplantation of primary tumor, tumor cells were s.c. transplanted to seven to 12-wk-old female NSG mice. 10^4–10^2 MM^+, $CD271^+$, or $MM^-/CD271^-$ cells were isolated by FACS from xenograft tumors, retransplanted in NSG mice and monitored for 10 wk. For short-term cultured cells, 10^3 MM^+ or MM^- cells were isolated by FACS from xenograft STC1 tumors and retransplanted or 10^4 MM^+ or $CD271^+$ cells were isolated by FACS from STC2 and transplanted into NSG mice (two inoculations in each group) and monitored for up to 15 wk. Before retransplantation, any dead cells were eliminated during sorting by PI-positivity and mouse cells were excluded by use of a mouse cell marker, H2Kd and/or human specific EpCAM. Tumorigenicity and the effect of GCNT1 silencing were assessed in 8-wk-old BALB/c nu/nu female athymic mice. In general tumor take was low in nu/nu mice compared with NSG mice. A total 1×10^7 of MM^+ or $CD271^+$ clones of MCF7 in suspension in PBS (150 μL) was s.c. transplanted at the fourth mammary gland with six inoculations in each group. 1×10^7 shGCNT1 or shControl (scrambled shRNA) treated MM^+ cells were tested with three inoculations in each group. Formation of tumors was assessed by palpation during the course of experiments (up to 15 wk).

RNA Isolation and mRNA/miRNA Expression Analysis. Total RNA was isolated using TRIzol Reagent (Invitrogen) and mRNAs of *ΔNP63*, *SNAI2*, *TWIST*, *FN*, and *GAPDH* were quantified using SYBR based qRT-PCR (iCycler, Bio-Rad) as previously described (5). PCR conditions were the following: 3 min at 95 °C, followed by 40 cycles of 95°C for 15 s and 60°C (55°C for POU5F1) at 45 s, followed by melting curve analysis. Additional primers that were used to quantify mRNA expression are following (F, forward; R, reverse):

SOX2 F, ACAGCGCCCGCATGTACAACA;
SOX2 R, ACGCGGTCCGGGCTGTTTTT;
POU5F1 F, AGTGAGAGGCAACCTGGAGA;
POU5F1 R, CAAAAACCCTGGCACAAACT;
NANOG F, CAACTGGCCGAAGAATAGCA;
NANOG R, GCAGGAGAATTTGGCTGGAA;
JAG1 F, CTGTAAGAATCTCATTGCCAGCTA ;
JAG1 R, ACACAGACACTGGAATCTGTTGAT ;
KLF5 F, CCACCACCCTGCCAGTTAAC;
KLF5 R, TAAACTTTTGTGCAACCAGGGTAA;
CD44 F, GTCCCATACCACTCATGGATCT;
CD44 R, CTTCTTCGACTGTTGACTGCAA;
VIM F, CGAAAACACCCTGCAATCTT ;
VIM R, TTGGCAGCCACACTTTCATA;
CDH1 F, AGTGCCAACTGGACCATTCA;
CDH1 R, TCTTTGACCACCGCTCTCCT;
GCNT1 F, GATGTCACCTGGAATCAGCA;
GCNT1 R, GCAGCAACGTCCTCAGCAT.

Each mean of expression of *SOX2*, *POU5F1*, *NANOG*, *JAG1*, *KLF5*, *ΔNp63*, *CD44*, *TWIST*, *SNAI2*, *FN*, *VIM*, and *CDH1* in triplicate was calculated by the ΔΔCt method, using *GAPDH* as a normalized control and represented in a logarithmic scale. Mature miRNAs of hsa-miR-205, hsa-miR-221, hsa-miR-222 were quantified based on the stem-loop qRT-PCR, by using the Taqman MicroRNA assays, according to the manufacturer's protocol (Applied Biosystems). Data were presented in a logarithmic scale as the mean expression of two different experiments performed in triplicate, normalized to the endogenous RNU48 expression by the ΔΔCt method.

Microarray Experiments. For gene expression microarray analyses, samples included three cloned $CD271^+$ cells (two different clones and FACS-sorted $CD271^+$ from a clone) and four cloned MM^+ cells (three different clones and FACS-sorted MM^+ from a clone) of MCF7. RNA quality was confirmed using Bio-Analyzer 2100 (Agilent Technologies). 400 ng total RNA was amplified and labeled using Agilent Low RNA Input Fluorescent Linear Amplification Kit and 1.65 μg of Cy3-labeled cRNA was hybridized onto Agilent Whole Human Genome Microarrays (4 × 44k, G4112F). For the miRNA analyses, 100 ng total RNA was labeled using the miRNA Microarray System (Agilent); two replicates of both a $CD271^+$ clone and a MM^+ clone were hybridized onto Agilent Human miRNA Microarrays (G4470B) containing 723 human and 76 human viral miRNAs. Array-based CGH was performed using Agilent Human Genome CGH 244K Microarrays according to the manufacturer's protocol (1 μg of DNA was used as input). All arrays were scanned on an Agilent Microarray Scanner G2565A.

GCNT1 Silencing. We used pGIPZ (RHS4330, Openbiosystems) lentiviral vector expressing a short hairpin RNA (shRNAmir) to silence *GCNT1* expression (mature sense sequence: GACACCTGACGACTAT ATA) or a scrambled nontargeted shRNA (RHS 4346, Openbiosystems) as a negative control. Lentiviral production and cell transduction was previously described (5). Puromycin (2 μg/mL, Sigma) was used to select for stable cell lines. Cell lines with a transduction rate over 80% were used for further studies. The efficiency of the knockdown was confirmed by qRT-PCR with GCNT1 primers which mentioned above, normalized by GAPDH. Levels of sialyated MUC1 expression compared with MM expression in GCNT1 silenced and control populations were assessed by 115D8 by FACS (1:100, Biogenesis) and immunostaining (1:100), as described above.

Data Analysis and Statistics. Microarray gene expression data were preprocessed using Agilent Feature Extraction Software (v 9.5.3.1) using default settings. Furthermore, data were quantile normalized and filtered for spot quality using GeneSpring GX Software (Agilent) leaving 30705 probes for analysis. For average-linkage hierarchical clustering analysis using Pearson distance measures, the genes were mean-centered. Cluster analyses were performed using Cluster (v 3.0) and displayed using Java Tree View (v 1.1.5r2). Genes significantly differentially expressed between MM^+ and $CD271^+$ cells were identified using two-class, unpaired SAM with < 2.5% false discovery rate (FDR) (8) and 1025 genes were confirmed by differential expression distance synthesis (DEDS) (9). The clustering and calculations for Fig. S5 were programmed in MATLAB. Fold change analysis of the miRNA microarray data used absolute ratio of the normalized intensities (no log scale) of two replicate experiments of each sample. For the CGH analysis, the raw data were extracted and normalized using Feature Extraction (v. 9.1.3.1, Agilent) and further processed and analyzed using CGH Explorer (http://www.ifi.uio.no/forskning/grupper/bioinf/Papers/CGH/). Survival analysis was performed using the R package (10) survfit for Kaplan–Meier analysis with log-rank test. For the MicMa data (11), breast cancer specific death was used as endpoint; UNC data (12), relapse-free survival; Miller data (13), disease-specific survival. Mapping of probes between the three different breast tumor datasets (Agilent and Affymetrix platforms) was performed using the R package biomaRt (14) and numbers are illustrated in Fig. S5B.

Nonparametric Mann–Whitney tests were performed to measure the differences in mammosphere-forming capacity, tumor formation and invasiveness between $CD271^+$ and MM^+ cells, by using R (v 2.13.0).

1. Rønnov-Jessen L, Van Deurs B, Nielsen M, Petersen OW (1992) Identification, paracrine generation, and possible function of human breast carcinoma myofibroblasts in culture. *In Vitro Cell Dev Biol* 28A:273–283.
2. Petersen OW, et al. (1990) Differential tumorigenicity of two autologous human breast carcinoma cell lines, HMT-3909S1 and HMT-3909S8, established in serum-free medium. *Cancer Res* 50:1257–1270.
3. Johannsson OT, et al. (2003) Characterization of a novel breast carcinoma xenograft and cell line derived from a BRCA1 germ-line mutation carrier. *Lab Invest* 83:387–396.
4. Villadsen R, et al. (2007) Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol* 177:87–101.
5. de Neergaard M, et al. (2010) Epithelial-stromal interaction 1 (EPSTI1) substitutes for peritumoral fibroblasts in the tumor microenvironment. *Am J Pathol* 176:1229–1240.
6. Gordon MN, Osterburg HH, May PC, Finch CE (1986) Effective oral administration of 17 beta-estradiol to female C57BL/6J mice through the drinking water. *Biol Reprod* 35:1088–1095.
7. Tomayko MM, Reynolds CP (1989) Determination of subcutaneous tumor size in athymic (nude) mice. *Cancer Chemother Pharmacol* 24:148–154.
8. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98:5116–5121.
9. Yang YH, Xiao Y, Segal MR (2005) Identifying differentially expressed genes from microarray experiments via statistic synthesis. *Bioinformatics* 21:1084–1093.
10. Bøvelstad HM, et al. (2007) Predicting survival from microarray data—a comparative study. *Bioinformatics* 23:2080–2087.
11. Enerly E, et al. (2011) miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS ONE* 6:e16915.
12. Prat A, et al. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12:R68.
13. Miller LD, et al. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 102:13550–13555.
14. Durinck S, et al. (2005) BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* 21:3439–3440.

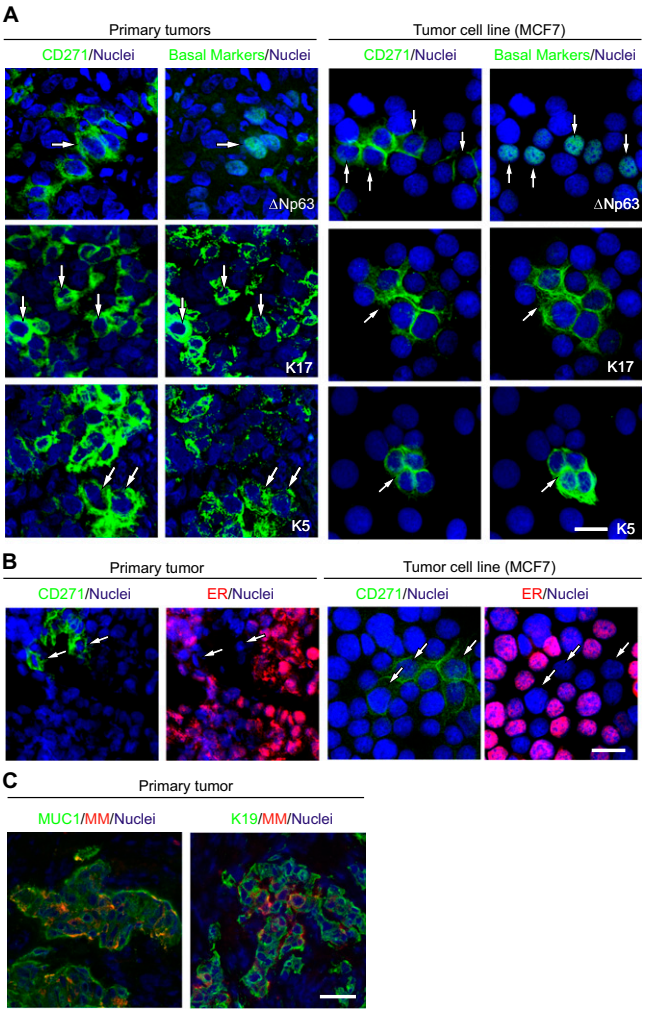

Fig. S1. CD271 and MM, two distinct differentiation markers of normal human breast identify distinct subsets of cells in primary tumors and cell lines. (*A*) CD271 in tumors and in tumor cell lines costains with basal-like lineage markers: Multicolor imaging of tumors (*Left*) and tumor cell line (*Right*) stained with CD271 (*left columns*) and basal-like markers ΔNp63, keratin K17, and keratin K5 (*right columns*) on a background of nuclear staining (blue). Note that the two columns under both tumors and tumor cell line headings indicate duel staining of the same field. The intention is always to show basal lineage affiliation in green and luminal affiliation in red (*A* and *B*). CD271 staining and staining with other basal markers shown overlap in the same cellular compartments (arrows). (Scale bar, 25 μm.) (*B*) ER negative cells are present in the CD271$^+$ compartment. Multicolor imaging of a tumor (*Left*) and a tumor cell line (*Right*) stained with CD271 (green) and ER (red) on a background of nuclear staining (blue). Note the mutually exclusive staining of CD271 (arrows) and the luminal lineage marker ER (arrows). (Scale bar, 25 μm.) (*C*) MM is part of a MUC1$^+$/ K19$^+$ compartment. Multicolor imaging of cryostat sections stained with MM (red) and MUC1 (green; *Left*) or keratin K19 (green; *Right*). Note that MM is confined to the MUC1$^+$/ K19$^+$ compartment. (Scale bar, 50 μm.)

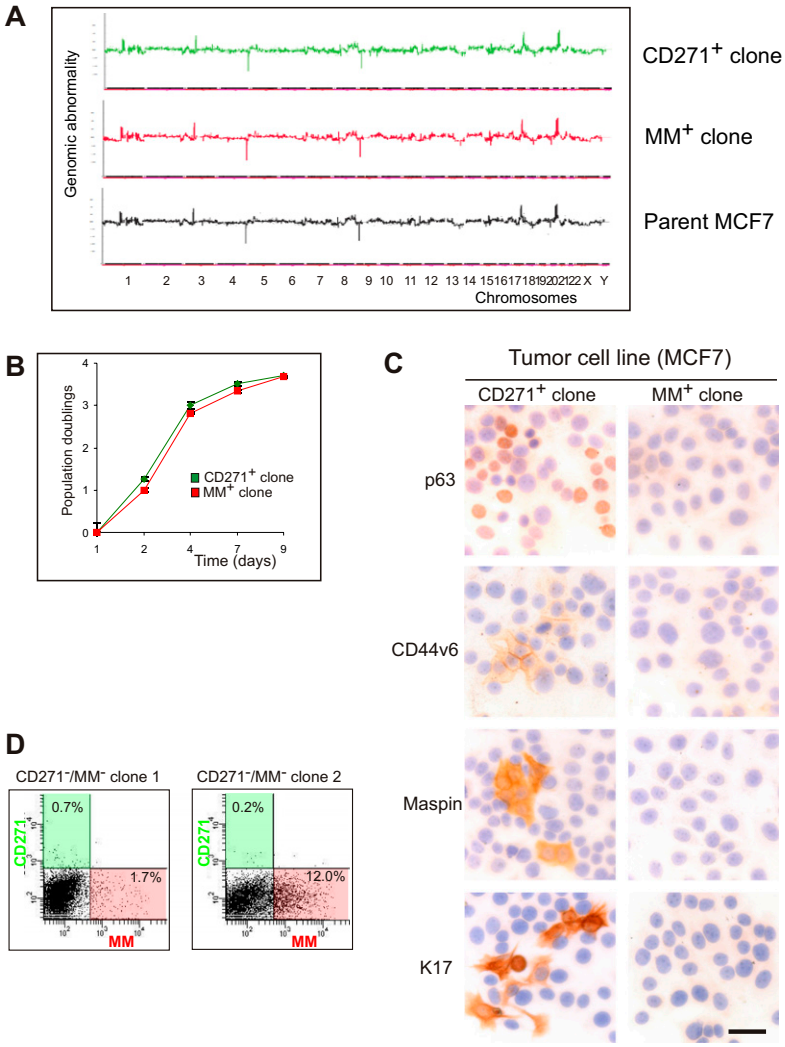

Fig. S2. CD271$^+$ -derived clones contain MM$^+$ populations, but freshly isolated MM$^+$ clones do not contain CD271$^+$. (*A*) CD271$^+$ and MM$^+$ clones of MCF7 have the same CGH signature as parent MCF7 population. Genomic profiling by array-based comparative genomic hybridization (CGH) show that the CD271$^+$ and MM$^+$ clones are indeed related with each other as well as with the initial population. Single plot graphical view along the chromosomes using the piecewise constant fit method for each of the three individual cell populations in the CGH Explorer program. Legend indicates logarithmic 2 CGH ratios on *y* axis. (*B*) CD271$^+$ and MM$^+$ clones have similar population doubling rates. Data represent the mean of cell numbers harvested from wells in duplicate at defined time points in a logarithmic 2 scale. (*C*) CD271$^+$ clones contain many basal markers, whereas MM$^+$ clones do not. Immunoperoxidase staining of the CD271$^+$ and MM$^+$ clones of MCF7. The basal markers p63, CD44v6, Maspin and K17 were positive in the CD271$^+$ clone (*Left*), and negative in the MM$^+$ clone (*Right*). Cells were counterstained with hematoxylin (blue). (Scale bar, 50 μm.) (*D*) CD271$^-$/MM$^-$ clones of a tumor cell line (MCF7) are less predictable in terms of CD271 and MM phenotype as demonstrated by the variation between the FACS profiles of two representative clones.

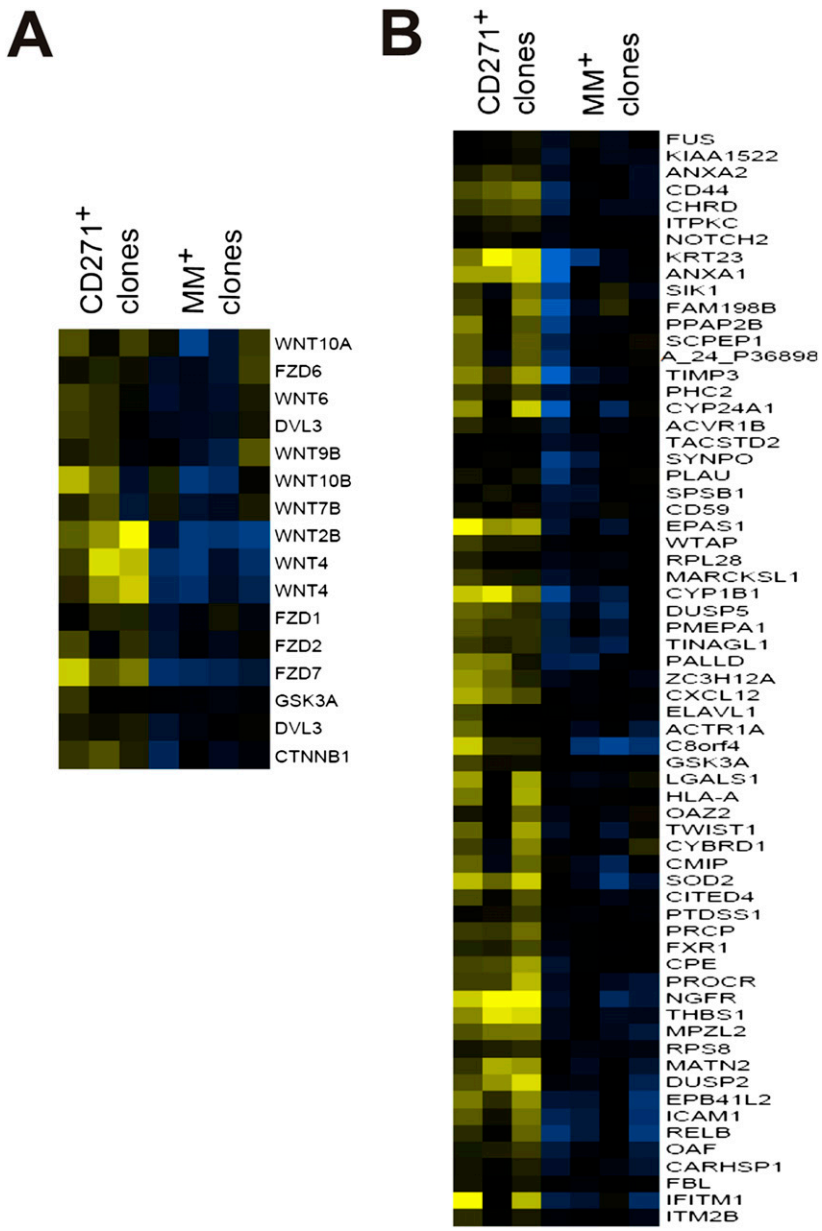

Fig. S3. CD271$^+$ and MM$^+$ clones are distinct populations by gene expression. (*A*) Members of the canonical Wnt signaling pathway are preferentially up-regulated in CD271$^+$ clones of MCF7. Heat map illustrating high expression of genes in the Wnt pathway in CD271$^+$ clones compared with MM$^+$ clones of MCF7. (*B*) A CD44$^+$ gene set is up-regulated in the CD271$^+$ clones of MCF7. Heat map illustrating high expression of genes found to be up-regulated in CD44$^+$/PROCR$^+$ cells from Shipitsin et al. (9) in CD271$^+$ clones compared with MM$^+$ clones from the MCF7 cell line. Average-linkage hierarchical cluster of genes and samples: Yellow = highest expression; blue = lowest expression, Black = average expression; gray = missing data.

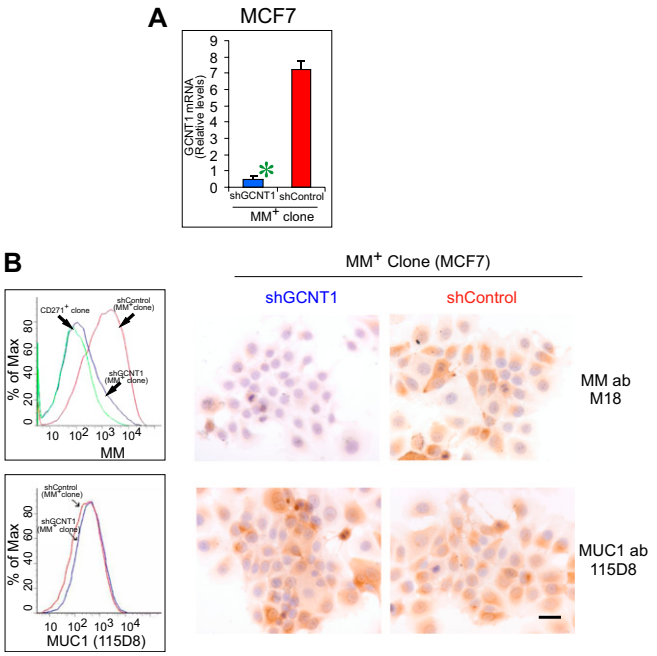

Fig. S4. Luminal-like cells within the hierarchy are tumor initiating and invasive in absence of basal-like traits. (*A*) shGCNT1 reduces GCNT1 mRNA expression. shGCNT1 or nontarget shRNA (shControl) was delivered into a MM$^+$ clone of MCF7, using lentivirus. *GCNT1* mRNA levels in shGCNT1 (blue bar) and control (red bar) were examined by qRT-PCR, normalized by GAPDH expression. Green asterisk indicates the relative GCNT1 level in a CD271$^+$ clone of MCF7. (*B*) GCNT1 silencing reduces MM terminal glycan expression. FACS diagram (*Left*) and immunoperoxidase staining (*Right*), showing the intensity of staining with MM (*Upper*) or MUC1 (*Lower*) in GCNT1 knockdown cells compared with control cells. Note the remarkable reduction in staining as a consequence of *GCNT1* silencing and the lack of impact of this on general MUC1 expression.

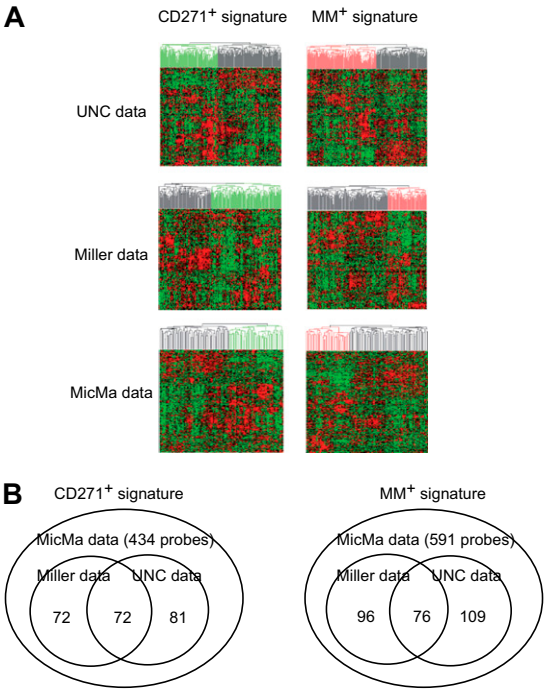

Fig. S5. CD271$^+$ and MM$^+$ gene signatures predict poor clinical outcome in breast cancer. (*A*) Gene expression pattern of tumor samples by hierarchical clustering. Hierarchical clustering of gene expression data from three independent breast cancer data sets, MicMa, UNC and Miller, using genes differentially expressed between CD271$^+$ and MM$^+$ clones. Two main groups of patients were identified in each data set for each signature and subjected to Kaplan-Meier analysis. Red colored dendrogram indicates tumors associated with poor survival by the MM$^+$ signature; green colored dendrogram indicates tumors associated with poor survival by the CD271$^+$ signature. (*B*) Venn diagrams illustrating the number of probes available for each of the two gene signatures, CD271 and MM in three independent gene expression datasets.

Table S1. Summary of characteristics of breast tumors analyzed

Tumor no.	ER	PR	K17	K5	ErbB2	CD271	MM	Experiments
1	++	++	—	—	Low	++	++	Imaging, Invasion (PT1)
2	—	—	++	++	High	+	+	Imaging (PT2)
3	—	—	++	++	Low	++	++	Imaging (PT3)
4	++	—	++	++	Low	++	+	
5	++	++	—	—	Low	+	+	
6	—	—	+	++	High	+	+	
7	++	++	—	—	Low	+	+	Invasion (PT7)
8	++	++	—	—	Low	++	++	
9	++	++	—	—	Low	++	+	
10	++	++	—	—	Low	+	++	
11	++	++	—	—	Low	+	++	
12	+	—	++	—	Low	+	+	
13	—	—	++	++	Low	++	+	
14	++	++	—	+	Low	++	++	
15	—	—	++	++	Low	+	+	
16	—	—	++	++	Low	+	++	
17	++	++	—	—	Low	++	++	
18	++	++	—	—	Low	+	++	
19	++	++	—	—	Low	+	++	
20	—	—	++	++	Low	+	++	
21	—	—	++	+	Low	—	++	
22	++	—	—	—	High	—	+	
23	—	—	—	—	High	—	++	
24	—	—	+	—	Low	—	++	
25	—	—	—	—	High	—	++	
26	—	—	—	—	Low	—	++	
27	++	++	—	—	Low	—	++	
28	++	—	—	—	Low	—	++	
29	++	++	—	—	Low	—	++	
30	—	—	++	++	Low	—	++	
31	++	++	—	—	Low	—	+	
32	++	++	—	—	Low	—	++	
33	++	++	—	—	Low	—	++	
34	—	—	++	++	Low	—	+	
35	++	++	—	—	Low	++	—	
36	++	++	++	+	High	+	—	
37	++	++	—	—	Low	++	—	
38	++	++	—	—	Low	+	—	
39	—	—	++	+	Low	++	—	
40	—	—	—		High	++	—	
41	++	++	—	—	Low	+	—	
42	++	++	—	—	Low	+	—	
43	++	—	—	—	Low	—	—	
44	++	—	—	—	Low	—	—	
45	—	—	++	++	Low	—	—	
46	++	++	—	—	Low	—	—	
47	++	++	—	—	Low	—	—	
48	++	++	—	—	Low	—	—	
49	++	—	—	—	Low	—	—	
50	++	++	—	—	Low	—	—	
51	++	—	—	—	Low	—	—	
52	—	—	—	—	Low	—	—	
53	++	—	—	—	Low	—	—	

Cryostat sections of 53 primary breast tumor specimens were stained by immunoperoxidase. Tumors were classified as $ER^{+}/PR^{+}/ErbB2^{lo}$, $ER^{\pm}/PR^{-}/ErbB2^{hi}$, and $ER^{-}/PR^{-}/ErbB2^{lo}$ based on staining also with K17 and K5. Primary tumor biopsies that were used for cell culture are indicated in the Experiments column. Note that biopsies analyzed were not selected by subtypes: All biopsies acquired are shown and those with a sufficient amount of CD271, EpCAM, and MM staining to allow analysis were used. —, no cells stained (stromal cells not included); +, any cancer cells stained; ++, two or more foci of cancer cells stained; low, no or faint staining in the majority of cells; high, strong staining in the majority of cells.

Table S2. Characteristics of primary and short-term cultured breast cancer samples for in vivo tumor formation

Sample	ER	PR	K17	K5	ErbB2	CD271	MM	Experiments	Tumor formation in mice
PT8	++	++	—	—	Low	+	++	Xenograft	No
PT9	++	—	—	—	Low	+	+	Xenograft, mammosphere, invasion	Yes
PT10	++	++	—	—	Low	+	+	Xenograft, mammopshere, invasion	No
PT11	—	—	—	—	High	+	++	Xenograft, mammopshere, invasion	No
PT12	—	—	+	—	Low	+	++	Xenograft	No
PT13	—	—	++	++	Low	++	+	Xenograft	Yes
PT14	—	—	++	++	Low	+	++	Xenograft	No
STC1 (HMT3909)	—	—	++	—	Low	++	++	Xenograft	Yes
STC2 (L56Br-C1)	—	—	++	++	Low	+	++	Xenograft	Yes

Nine primary breast cancer samples, including 7 uncultured (PT8-14), which were selected from a repository of 28 additional samples based on positive staining for both CD271 and MM, and two short-term cultured primary carcinomas (STC1-2) were used for in vivo tumor-initiating in NSG mice. Three samples (PT13, STC1, and STC2) were successfully tumor initiating and used for further characterization of $CD271^+$ and MM^+ cell behavior. —, no cells stained (stromal cells not included); +, any cancer cells stained; ++, two or more foci of cancer cells stained; low, no or faint staining in the majority of cells; high, strong staining in the majority of cells.

Dataset S1. Differentially expressed probes between $CD271^+$ and MM^+ by SAM analysis

Dataset S1 (XLS)

(*A*) List of probes and genes in the CD271 signature. (*B*) List of probes and genes in the MM signature.